
Proceedings of FIKUSZ '09 Symposium for Young Researchers, 2009, 195-203 © The Author(s). Conference Proceedings compilation © Budapest Tech Keleti Károly Faculty of Economics 2009. Published by Budapest Tech Keleti Károly Faculty of Economics, Tavaszmező u. 15-17. H-1084 Budapest, Hungary. http://kgk.bmf.hu/fikusz

Better data model makes less work?

Péter SZIKORA

Budapest Tech Keleti Károly Faculty of Economics, Institute for Organisation and Management

szikora.peter@kgk.bmf.hu

My goal is to prove a better data model can help us to work less. If we plan before work, we can save our resources, such as time, money, human resources and work. I would like to prove it with solving a reincurring problem: In Hungary students finishing their studies need to pass the final exam, which is organised twice a year, in each exam period. Some years ago we filed the reports by pen, nowadays we use Microsoft Office (Word, Excel), but either way, administration remained a largely mechanical, tedious task. I developed a data model, and some software to automate the task. With this solution the administrators have less work in the exam period. They can save about 8 hours per exam. The solution I propose uses free softwares[1]. Not only because it is free of charge, but first of all because in this case I have the freedom to change even the program codes, and anybody can download and use them without any administrative tasks. All one would need: one computer running a web server with MySQL and PHP, a printer and my software.

Keywords: database efficiency, data model, free software
JEL Code: L86

1 The problem

This paper describes a method to save our resources by solving an incurring problem, which is well known for all those interested in organizing and executing final examinations. What do we mean by resources? Most importantly: work. One would like to leave the work to the computer. For the aforementioned problem a well-designed database and some simple php-code can be the remedy. The second important resource is time. This is related to work. If we need to work less, we need less time to do that. The third important resource is money. That is, using free softwares of the same quality instead of buying expensive ones, or writing simple programs to solve a task instead of purchasing a for-fee program do that. Last but not least security is a very important resource as well.

Speaking about free softwares, 'free' means freedom, the freedom of using, understanding, modifying, developing and/or (re)distributing of the 'free' software. In addition, free softwares are usually free of charge, too, but this is not part of the

1 FSF.org, 2009. What is Free software?, http://www.fsf.org/about/what-is-free-software

definition of free softwares. According to the above softwares can have a better quality than proprietary, commercial softwares because there are no business aspects (i.e. to achieve profit) while developing them. Free softwares usually have much bigger testing teams than enterprises can even dream of. In case of using free software there is no dependency: if the original developers/maintainers of the software stop their work, you can and are allowed to continue. You can utilize the software in any way to have a more proper tool for your purpose. And last but not least, by using free softwares one should reach a higher level of computer security because of having their source code, too, while in case of commercial softwares you have only binaries without any source codes and in addition you are explicitly forbidden even to try to check the software you use.

There is no widely accepted solution for this problem in other universities. E.g. In Eötvös Loránd University, in the Faculty of Informatics they use the old method, they print all the report before the exam, and after they put the marks, questions and other data to the paper by hand. In other universities the old method, or the mail merge function is used.

2 The administration of final exams

Final exam is the last exam before students in higher education finishing their studies. It consists of three or more parts. First one is the thesis defence followed by oral exams from each of the two or more subjects of the final exam. The court reporter (secretary) needs to make a detailed report about it. In the past this report has been filed by hand. The administration of the final exam is now simple.

We got all the paper printed before the exam, and at the exam we wrote down the data on the paper by hand. The first innovation was printing the needed documents after the exam, using Microsoft Excel and Word's mail merge function[2].

It is a very simple system. We need to do the work before the final exam, make the spreadsheet, check the links and the Vlookup function. Vlookup is a built-in function in Excel, searches for a given value in the first column of a table array and returns a value in the same row of another column in the table array[3]

I've got spreadsheet file before the final exam from the department of study with the data of the students. I need to put together all other data, e.g. thesis title, questions from the opponent of the thesis, data of the committee of final exam (number, participants, date, venue).

When the final exam starts all participants of the committee got a paper, they can write there the students' mark. When the exam is over, I enter the results into the previously prepared spreadsheet and generate a word document by mail merge. After

2 Microsoft.com, 2009. Microsoft Mail merge Function, http://office.microsoft.com/en-us/word/ch060832701033.aspx

3 Microsoft.com, 2009. Excel Vlookup Function, http://office.microsoft.com/en-us/excel/HP052093351033.aspx.

that I only need to print it for all of the participants of the committee and they sign it. The results to be entered are only five number per student (three marks and two question numbers).

It is a simple but not the best system. There are some problems with it. The first problem is, if the column number is not the same as in the last case (e.g. we put a new column or remove an old column from the spreadsheet), the Vlookup function will not work properly. The second problem is that this procedure is very slow. I need more than 10 minutes after the exam, and much more time before it to prepare.

Last but not least, it needs too much work. I work in IT, and I like when computer works instead of me. (I worked in some company writing php scripts, and I thought I could make a little database, and some simple php code to solve this problem.)

A few months ago we developed a little, clear and consistent data model for testing a SIS module.[4] We would like to prove in that presentation the same as this work, that better data model saves our resources. Our goal was that we can make a better system than the existing Scholar Information System (SIS). We used that data model and free softwares and a normal pc as test environment. This application is described more precisely in my paper, what I wrote to 2nd International Conference for Theory and Practice in Education in May 2009.[5]

The results of that experiment were better than we had expected before. We can make a paper from this results in 7th International Conference on Management, Enterprise and Benchmarking in June 2009[6],[7] The result was the next: All the registration of all the pupils were successful.

3 The background

I need a server running Linux operating system and in addition to that runs a database management system (MySQL), a web server (Apache), and PHP as a host programming language. This server machine is a normal pc computer running free softwares only. I use another computer (a notebook) at the place of the final exam. All court secretaries of our institute at the final exams use notebooks to record all the data. These notebooks have internet connections thus they can connect to the server with http protocol.

I need a printer connected to the server. But if I want I can connect a printer to the notebook instead of the server.

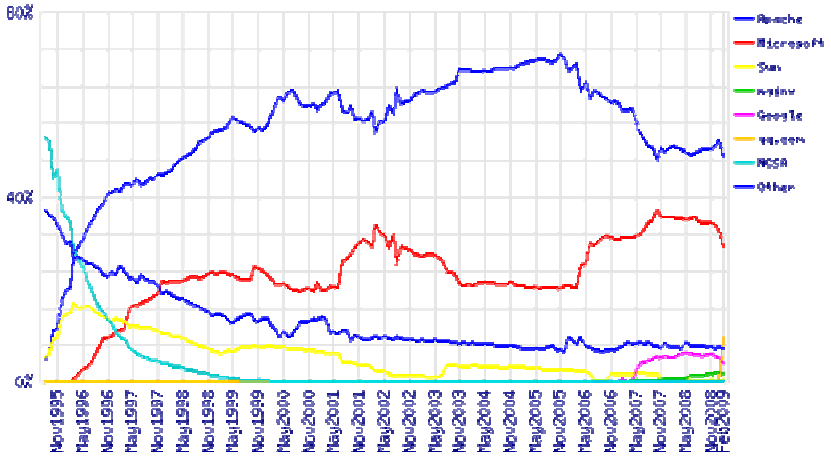
I have chosen Linux Operating system because it is free and responsible. Linux is a free Unix like operating system originally created by Linus Torvalds with the

4 Keszthelyi András, 2009.
5 Szikora Péter, 2009.
6 Szikora Péter, 2009.
7 Keszthelyi András, 2009.

assistance of developers all around the world. Developed under the GNU General Public License the source code for Linux is freely available to everyone.

Figure 1

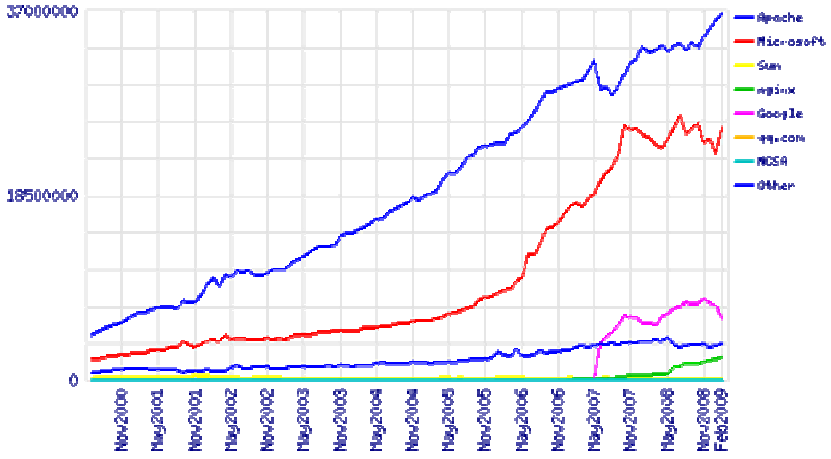
Market Share for Top Servers Across All Domains August 1995 - February 2009 [8]



The reasons for choosing Apache as the web server are numerous. The primary reason, however, to boil down the licensing, and the active support and development. Apache, on top of Linux, is used in most cases in all over the world so it can be considered the most accurately tested http server. See Figure 1 and 2.

Figure 2

Totals for Active Servers Across All Domains June 2000 - February 2009 [9]



8 Netcraft.com, 2009. Apache, http://news.netcraft.com/archives/2009/02/18/february_2009_web_server_survey.html

9 Netcraft.com, 2009. Apache, http://news.netcraft.com/archives/2009/02/18/february_2009_web_server_survey.html

The MySQL database management system has become the world's most popular open source DBMS because of its consistent fast performance, high relability and the ease of use. MySQL runs on more than 20 platforms including Linux, Windows.

Top reasons to use MySQL: scalability, flexibility, high performance and availability, strong data protection, open source freedom and 24 x 7 support and lowest total cost of ownership.[10]

Figure 3

Usage Stats for April 2007 [11]

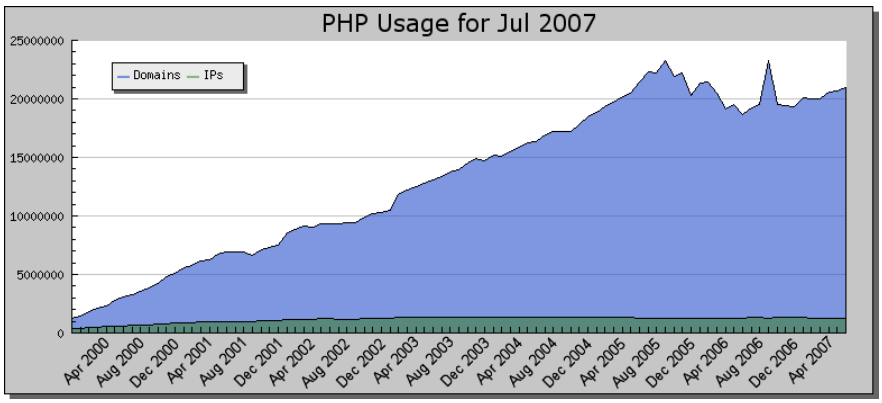


Figure 3 shows that PHP brake the 20 million domain border, which means that more than 20 million sites running PHP. PHP has the best documentations and it is free, too, and very easy to use.

Last but not least we need a free PDF[12] reader, too, e.g. Adobe Reader. Adobe Acrobat is a family of computer programs developed by Adobe Systems, designed to view, create, manipulate and manage files in Adobe's Portable Document Format (PDF)[13].

4 Data model

What can we call the efficiency of a database? It's capability to cope with high loads. This capability is determined by some very different factors: hardware environment, software environment (operating system, relational database management system, application programming language and tools, application programs themselves) and last, but not least, the quality of the data model[14].

In my case the influence of the hardware environment is not very important, because this program will be used by 2-3 people only. The most important element is the

10 Mysql.com, 2009. Mysql Documentation, http://dev.mysql.com/doc/, 2009
11 PHP.net, PHP usage, http://www.php.net/usage.php
12 Wikipedia, PDF, http://en.wikipedia.org/wiki/Portable_Document_Format, 2009.
13 Wikipedia, Adobe Acrobat, - http://en.wikipedia.org/wiki/Adobe_Acrobat, 2009
14 Keszthelyi András, under publication 2009.

software, the operating system, the relational database management system and the application itself. Choosing the programming language and tools there are two main fields which determine the performance of the developed program. These are the quality of the applied algorithms and the quality of program coding.

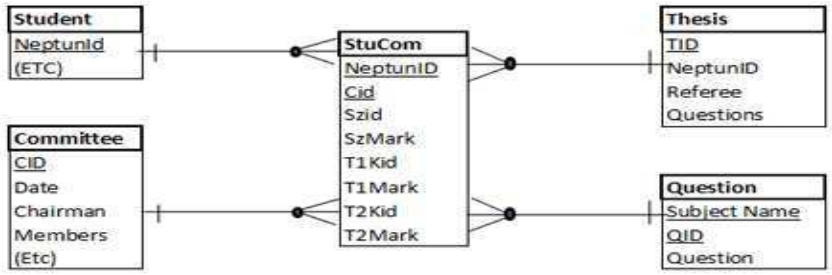
The main steps in developing an information system are: determining what we want as exactly as possible, data modelling (i.e. determining the data structure), determining the functions operating on the data structure. In case of data-intensive systems the data structure is more important and determines the functionality[15].

So data modelling is the basis, which is necessary but of course not sufficient for succeeding. The basis is only a possibility on which a good information system can be constructed. In order to succeed we need three-level data modelling and planning according to dr. Halassy[16]. These levels are the conceptual, the logical and the physical levels.

So after hardware and software has been chosen I need to develop a good data model. There are general prerequisites of the goodness of data models. At the conceptual level a good data model needs to be understandable, unambiguous, realistic, full and *minimal*[17]. In my data model there are 5 entities. You can see the data structure in Figure 4.

Figure 4

Conceptual data model schema



In the Student table the primary key is the NeptunID. The NeptunId is an unique identifier, generated by Neptun Scholar Information System18. The other fields are e.g. birth place, birth date, and some personal data. In the Committee table there are some data about the committees e.g. date, name of the chairman, members of the committee, venue, and some other data. In this table the primary key is the Committee ID. In the Thesis table, there are the titles of the thesises, the names of the referees and the questions for thesis defence. The primary key is the Thesis ID., The fourth table is the Question table. In this table there are the questions for final exam's

15 Raffai Mária dr.: Developing and Manage Informational systems (Információrendszerek fejlesztése és menedzselése). p. 541. Novadat Bt., 2003

16 Halassy Béla dr., 2002. Data modelling (Adatmodellezés). pp. 28-33. Nemzeti Tankönyvkiadó Rt.,

17 Halassy Béla dr., 1995. The fundamentals and secrets of Database design (Az adatbázistervezés alapjai és titkai). p. 192. IDG Magyarországi Lapkiadó Kft.,

18 SDA Stúdió Kft, 2009. Neptun.Net http://www.sdakft.hu/Letolthetok/Neptun_tajekoztato_anyag.pdf

subjects. The primary key is a compound key, it consists of the subject Name, and Question ID. The main table is the StuCom, This table connects the students and the committees to each other and contains the ordinal number of the questions students had to answer and their marks. In these tables we can store all the data needed for the final exam report. The primary key is a compound key, it consists of the fields NeptunID, and CommitteeID. The foreign keys are: Thesis ID, QuestionId and Committee ID. At this point I only need a good and simple program to archive my goal.

5 PHP Code

Once I made the data model, I need to develop the program. The program will have some modules. The first module is responsible for data import. I have two different excel files, first from the department of study and the second which contains all the data of the committee of final exam. I need to import it to my database. After it I can choose a committee and enter the question ordinal numbers and marks for the students.

All secretaries have access to their committee. They can change all data of their own committee, and if the data are correct, print it. So the second module is the data modifying one and the third module is the printing one. The whole system can be reached via HTTP protocol. If we want we can use it even without the network as well. All free softwares I use for this system are available for Windows and for Linux, too.

The program is developed in PHP[19]. PHP, which stands for "*PHP: Hypertext Preprocessor*" is a widely used open source general-purpose scripting language that is especially suited for web development and can be embedded into HTML. Its syntax draws upon C, Java, and Perl and is easy to learn except if somebody is socialized on Pascal. The main goal of the language is to allow web developers to write dynamically generated web pages quickly, but you can do much more with PHP. The PHP script needs to be integrated in short blocks, because of the developing and manageability. I use sessions in the code, because we can distinguish the users in a secure enough way by this. When I use a session, the visitor who is accessing the web site is assigned to a unique id, to the so-called session id. This is either stored in a cookie on the user side or is propagated in the URL[20]. The program itself is not finished yet, but I hope it would be for the next final exam period in January 2010.

6 Conclusions

By using this program we will save a lot of time. Do not need to integrate all existing excel files, only need to import to the system. And in the final exam, all secretaries will only need to put question numbers and marks into the system, and after the exam they can print all the required reports from the system in a few minutes. We save our resources. We need to work only 10 minutes in every final exam period for each final

19 Peter Moulding, 2002. PHP Black Book pp 9-12, Perfact-Pro Kft.
20 PHP.net, 2009. PHP Documentation, http://www.php.net/manual/en/

exam without using this program, we need to work more than 10 hours to get ready to final exam.

I hope we can use this system to simplify the administration of the final exams. I think if we make a good data model we need to work less. If it is true, I will reach my goal. I hope in the next final exam period we will see this program's advantages.

Acknowledgement

First of all, I would like to thank my colleague András Keszthelyi for all of his help in this work. Also I would like express my thanks to György Kadocsa who played a decisive role. Several people have expressed their willingness in this subject.

References

FSF.org, 2009. What is Free software?, http://www.fsf.org/about/what-is-free-software

GNU.org 2009. Free software, http://www.gnu.org/philosophy/free-sw.html,

Halassy Béla dr., 2002. Datamodelling (Adatmodellezés). Nemzeti Tankönyvkiadó Rt.

Halassy Béla dr., 1995. The fundamentals and secrets of Database design (Az adatbázistervezés alapjai és titkai). IDG Magyarországi Lapkiadó Kft.,

Keszthelyi András, 2009. How to Measure an Information System's Efficiency?, 7th International Conference on Management, Enterprise and Benchmarking, Budapest, Hungary, June 5-6

Keszthelyi András, under publication 2009. Remarks on the Efficiency of Information Systems, Acta Polytechnica Hungarica, Vol. 6. No. 3. pp.

Keszthelyi András, 2009. The Role of Data Modeling in Information System Efficiency. 2nd International Conference for Theory and Practice in Education, Budapest

Microsoft.com, 2009. Excel Vlookup Function, http://office.microsoft.com/en-us/excel/HP052093351033.aspx.,

Peter Moulding, 2002. PHP Black Book, Perfact-Pro Kft.

Mysql.com, 2009. Mysql Documentation, http://dev.mysql.com/doc/

Netcraft.com, 2009. February 2009 Web Server Survey http://news.netcraft.com/archives/2009/02/18/february_ 2009_web_server_survey.html

PHP.net, 2009. PHP Documentation, http://www.php.net/manual/en/,

PHP.net, 2009. PHP usage, http://www.php.net/usage.php,

Raffai Mária dr., 2003. Developing and Manage Informational systems (Információrendszerek fejlesztése és menedzselése). Novadat Bt.,

SDA Stúdió Kft, 2009. Neptun.Net http://www.sdakft.hu/Letolthetok/ /Neptun_tajekoztato_anyag.pdf

Szikora Péter, 2009. Measured Performance of an Information System. 7th International Conference on Management, Enterprise and Benchmarking, Budapest

Szikora Péter, 2009. The Role of the Tools and Methods of Implementation in Information System Efficiency. 2nd International Conference for Theory and Practice in Education, Budapest

Wikipedia, 2009. Portable Document Format, http://en.wikipedia.org/ /wiki/Portable_Document_Format

Wikipedia, 2009. Adobe Acrobat, - http://en.wikipedia.org/wiki/Adobe_Acrobat

Wikipedia, 2009. Open Source definition, http://en.wikipedia.org/wiki/Open_source